

Turbo Receiver Design for Phase Noise Mitigation in OFDM Systems

Gokul Sridharan and Teng Joon Lim Electrical and Computer Engineering

Department

University of Toronto, Canada

Email:gsridharan@comm.utoronto.ca, limtj@comm.utoronto.ca

Abstract

This paper addresses the issue of phase noise in OFDM systems. Phase noise (PHN) is a transceiver impairment resulting from the non-idealities of the local oscillator. We present a case for designing a turbo receiver for systems corrupted by phase noise by taking a closer look at the effects of the common phase error (CPE). Using an approximate probabilistic framework called variational inference (VI), we develop a soft-in soft-out (SISO) algorithm that generates posterior bit-level soft estimates while taking into account the effect of phase noise. The algorithm also provides an estimate of the phase noise sequence. Using this SISO algorithm, a turbo receiver is designed by passing soft information between the SISO detector and an outer forward error correcting (FEC) decoder that uses a soft decoding algorithm. It is shown that the turbo receiver achieves close to optimal performance.

I. INTRODUCTION

Increase in the demand for higher data rates has led to OFDM becoming the technology of choice for next generation wireless standards such as WiMAX and LTE. The deployment of the devices built around these standards brings to light the various implementation issues that become critical when designing an OFDM based system. PHN is an impairment that is quite different from the other impairments since it is a continuous time noise process that cannot be compensated for in the training stage. In this paper we consider the problem of data detection in an OFDM system that has been corrupted by PHN.

PHN arises from imperfections in the local oscillator (LO) that result in a random drift of the LO phase from its reference and is commonly modelled as a Wiener process. But, in general, a local oscillator is usually followed by a phase locked loop (PLL) that keeps the phase and frequency matched to a reference and in such a scenario PHN is better modelled as a wide sense stationary (WSS) process with bounded variance [1]. In this paper we assume phase noise to be a first order auto-regressive (AR(1)) process as suggested in [2] for the IEEE 802.11g standard. The effect of PHN has been studied extensively [3]–[6]. The effect of PHN can be split into two: the rotation of all the sub-carriers by a certain angle called the common phase error (CPE) and the leakage of the neighboring sub-carriers resulting in ICI. CPE is the average of the PHN sequence spanning an OFDM symbol.

PHN mitigation for single carrier systems and multi carrier systems such as OFDM has been studied extensively [5], [7]–[11]. While for single carrier systems it is possible to construct the factor graph and adopt a message passing algorithm [11] such an approach becomes prohibitively complicated for OFDM systems. Most algorithms suggested for PHN estimation in OFDM systems involve a decision directed process where the data symbols are estimated ignoring PHN and these decisions are fed back to estimate the PHN. This process is iterated either for a fixed number of times or until convergence. These are hard decision algorithms and do not provide any soft information on the decisions made. Another class of techniques that have been used to jointly estimate PHN and data symbols include the Markov chain Monte Carlo methods [8]. The major concern with decision directed algorithms is the assumption that the initial crude estimation of data symbols ignoring PHN results in a majority of the symbols being detected correctly while a good fraction of the remaining symbols get corrected over the subsequent iterations. As we will see in a subsequent section, this need not necessarily be true and can lead to a premature error floor. This motivates the need for a turbo receiver that exchanges messages between the symbol detector and the outer FEC decoder.

While [7] developed algorithms for PHN mitigation using VI that generated soft symbol estimates using a Gaussian approximation, we design a soft detection algorithm that computes soft bit estimates of the transmitted bits using the VI framework and a discrete distribution for the bits. The algorithm is capable of incorporating prior information on bits and hence fits in naturally with a soft decoding algorithm for the FEC code. The paper is organized as follows. Section II reviews modelling of phase noise, section III sets up the received signal model and

discusses the consequences of phase noise on the received signal. In section IV the VI framework for this problem is presented along with the bit level detection algorithm. The last section presents the simulation results.

II. PHASE NOISE CHARACTERISTICS

Since almost all communication systems employ an LO followed by a PLL, we assume PHN to be a WSS process. In particular, we model it as an AR(1) process. Such a model has been adopted by IEEE standards such as 802.11g where phase noise is modelled as the output of a single pole Butterworth filter driven by a zero mean white noise process. Such a process has an autocorrelation function of the form [7]

$$R_p(k) = \sigma_\theta^2 e^{-2\pi\Omega_o T_s |k|}. \quad (1)$$

Here, $1/T_s$ is the sampling rate and Ω_o is the one sided 3-dB bandwidth of the oscillator. σ_θ is the RMS value of the PHN process in radians, a typical value being around 3° . In general, the 3-dB bandwidth of an oscillator tends to be in the range of hundreds of kHz. One can write the covariance matrix of a length-N sequence of PHN as

$$\mathbf{\Phi} = \sigma_\theta^2 \begin{bmatrix} 1 & p & p^2 & \dots & p^{N-1} \\ p & 1 & p^3 & \dots & p^{N-2} \\ \vdots & \vdots & \ddots & \dots & \\ p^{N-1} & p^{N-2} & p^{N-3} & \dots & 1 \end{bmatrix} \quad (2)$$

where, $p = e^{-2\pi\Omega_o T_s}$. Usually, the cut off frequency of the Butterworth filter (Ω_o) is set to the 3-dB bandwidth of the oscillator.

A. Sample mean statistics

Since the CPE plays a critical role in the detection of symbols in affected by PHN, we take a closer look at the sample mean of a length-N PHN sequence. Suppose we let $\bar{\theta}$ denote the sample mean of a length-N sequence of the phase noise process, we have,

$$\bar{\theta} = \frac{1}{N} \sum_{k=1}^N \theta[k]. \quad (3)$$

Since the statistics of the phase noise process are assumed to be known, it can be shown that $\bar{\theta}$ is a zero mean Gaussian random variable with variance $\mathbf{1}^T \mathbf{\Phi} \mathbf{1} / N^2$ [12]. Note that the variance of the sample mean is a function of N and goes to zero as $N \rightarrow \infty$.

III. RECEIVED SIGNAL MODEL

A. Gray mapping : bits to M-QAM symbols (from [13])

We assume M to be a perfect square. Let L denote $\log_2 M$ and let $\mathbf{B} \in \{-1, 1\}^{N \times L}$ represent a matrix of $N \times L$ bits that need to be mapped to N M-QAM symbols. Denote the columns of \mathbf{B} as $\mathbf{b}_{r1}, \mathbf{b}_{r2}, \dots, \mathbf{b}_{rL/2}, \mathbf{b}_{i1}, \mathbf{b}_{i2} \dots \mathbf{b}_{iL/2}$. If the resulting vector of M-QAM symbols is represented as \mathbf{d} , then, from [13], we have the following relation :

$$\mathbf{d} = \sum_{l=1}^{L/2} 2^{l-1} \prod_{p=1}^{p=L/2} \mathbf{b}_p + j \sum_{l=L/2+1}^L 2^{l-1} \prod_{p=L/2+1}^{p=L} \mathbf{b}_p \quad (4)$$

Here, \prod represents element wise product of the vectors. Let the function f denote the mapping from bits to symbols. Thus, $\mathbf{d} = f(\mathbf{B})$. Further, denote the matrix $[\mathbf{b}_{r1} \mathbf{b}_{r2} \dots \mathbf{b}_{rL/2}]$ as \mathbf{B}_r , and the matrix $[\mathbf{b}_{i1} \mathbf{b}_{i2} \dots \mathbf{b}_{iL/2}]$ as \mathbf{B}_i and define the functions f_r and f_i as :

$$f_i(\mathbf{B}_i) = j \cdot \Im(f(\mathbf{B})) \quad (5)$$

$$f_r(\mathbf{B}_r) = \Re(f(\mathbf{B})) \quad (6)$$

Hence, $\mathbf{d} = f(\mathbf{B}) = f_r(\mathbf{B}_r) + f_i(\mathbf{B}_i)$.

B. The received signal

In this work, we consider the detection of an OFDM symbol transmitted over a block fading frequency selective channel, where the channel stays constant over the duration of one OFDM symbol. We also assume that perfect frame synchronization, including carrier frequency recovery have been established in the training stage. We further assume that current channel conditions have been estimated during the training phase and that channel state information is available on the receiver side. Algorithms that can estimate the channel in the presence of PHN and carrier frequency offset have been presented in [8], [12]. In the data detection stage we assume that the received symbol vector has been affected by PHN in addition to the channel and the additive noise. The received signal for such a scenario in the discrete domain after appropriate sampling and removal of the cyclic prefix is given by

$$\mathbf{r} = \mathbf{P}\mathbf{F}^H \mathbf{H} \mathbf{d} + \mathbf{n} = \mathbf{P}\mathbf{F}^H \mathbf{H} f(\mathbf{B}) + \mathbf{n}. \quad (7)$$

Here, \mathbf{F} is an $N \times N$ DFT matrix with the (l, m) th entry given by $\mathbf{F}_{lm} = (1/\sqrt{N})e^{-(2\pi j(l-1)(m-1)/N)}$, \mathbf{P} is the diagonal matrix given by $\text{diag}(e^{j\theta}) \approx \text{diag}(\mathbf{1} + j\theta)$, where θ is the PHN sequence and $\mathbf{H} = \text{diag}(\mathbf{h})$ is the channel matrix in the frequency domain and the $N \times L$ binary matrix \mathbf{B} contains the transmitted bit sequence (\mathbf{d} is the corresponding symbol sequence). \mathbf{n} is complex white Gaussian noise with variance σ^2 per dimension.

Detection schemes that ignore PHN involve computing the DFT of the received signal and then adopting MMSE or zero forcing detection on individual sub carriers. The DFT of the received signal in the presence of phase noise results in a vector $\mathbf{R} = [R_0 R_1 \dots R_{N-1}]^T$ which can be written as [7]

$$R_k = c_0 d_k h_k + \sum_{l=0, l \neq k}^{N-1} d_l h_l c_{(l-k) \bmod N} + \nu_k. \quad (8)$$

Here, the vector $\mathbf{c} = [c_0 c_1 \dots c_{N-1}]^T$ is given by $(1/\sqrt{N})\mathbf{F}\mathbf{p}$ (where $p = e^{j\theta}$) i.e. the frequency domain representation of the PHN sequence. It can be shown that ν is an uncorrelated white noise process with $\nu_k \sim \mathcal{CN}(0, 2\sigma^2)$. Eq.(8) clearly illustrates how PHN affects the received signal. Note that c_0 is the CPE i.e $1 + \frac{1}{N} \sum \theta_k$ (under small angle assumption), and its effect is to rotate every received symbol by the average phase angle $\bar{\theta}$. The next subsection further elaborates the consequences.

C. Effect of phase noise on the received signal and its consequences

Decision directed feedback mechanisms that have been suggested to counter the ICI resulting from PHN involve an initial step where the symbols are detected in a crude manner, ignoring the effects of PHN. The underlying assumption is that a majority of symbols detected in this way are correct and hence a back substitution of these symbols into the received signal structure must aid the detection process. This need not be true for certain realizations of the phase noise, specifically, scenarios where the CPE is high. This is because the received symbol gets rotated by an angle equal to the CPE and if this angle is greater than the average angular separation between the adjacent symbols of the constellation, the symbol is likely to be detected in error. As an example, for the 64-QAM constellation a rotation of 9° causes a symbol error with probability 0.43 in the absence of noise. When the oscillator linewidth is set to 100kHz and the RMS value of phase noise is set to 3° , the probability of encountering a length 64 (assume 64 sub-carriers in one OFDM symbol) phase noise sequence whose mean is greater than 9° is 4×10^{-5} . This

leads to a premature error floor.

One can counter this by either embedding pilots in the OFDM symbol and using them to estimate and compensate for CPE [5] or by designing a blind turbo receiver. While embedding pilots is a common practice to aid channel estimation, this comes at the cost of spectral efficiency. In scenarios where embedding pilots in every OFDM symbol is not an option one needs to consider designing a turbo receiver. Since the variance of the CPE decreases with increasing length of the PHN sequence, there exists a compelling case to design a turbo receiver for short length OFDM symbols. Having made the case for joint detection-decoding, we design a bit level detection algorithm that generates soft bit estimates and use it to set up a turbo receiver. The next subsection explains the basic principles involved.

D. Actual and postulated posterior distributions

From [7], we note that by approximating $e^{j\theta}$ to $\mathbf{1} + j\theta$ under the small angle assumption, one can write the conditional distribution of \mathbf{r} given \mathbf{B} and θ to be :

$$p(\mathbf{r}|\mathbf{B}, \theta) = \mathcal{CN}(\text{diag}(\mathbf{P}\mathbf{F}^H \mathbf{H}f(\mathbf{B}))(\mathbf{1} + j\theta), 2\sigma^2\mathbf{I}) \quad (9)$$

Further, in [7] it was shown that the optimal detector (i.e. ML estimate of \mathbf{d}) for such a signal model, given the prior distribution of the phase noise sequence θ has an exponential complexity in N . Given that the distribution $p(\mathbf{r}|\mathbf{B})$ and consequently the posterior distribution $p(\mathbf{B}|\mathbf{r})$ do not lend themselves to efficient optimal detection, we look at an approximation to $p(\mathbf{B}|\mathbf{r})$, such that, computing the optimal estimate corresponding to the approximated distribution is straightforward. To this end, we note that $p(\mathbf{B}|\mathbf{r})$ is computed by marginalizing $p(\mathbf{B}, \theta|\mathbf{r})$ with respect to θ , i.e.

$$p(\mathbf{B}|\mathbf{r}) = \int p(\mathbf{B}, \theta|\mathbf{r})d\theta \quad (10)$$

The variational inference approach approximates $p(\mathbf{B}, \theta|\mathbf{r})$ with a function $Q(\mathbf{B}, \theta)$ of the form $Q_{\mathbf{B}}(\mathbf{B})Q_{\theta}(\theta)$ such that

$$Q(\mathbf{B}, \theta) \sim p(\mathbf{B}, \theta|\mathbf{r}) \quad (11)$$

This kind of an approximation is equivalent to assuming that \mathbf{B} and θ are independent conditioned on \mathbf{r} and this in turn implies the maximizer of the distribution $Q_{\mathbf{B}}(\mathbf{B})$ is the optimal estimate

of \mathbf{B} . We further assume that the distribution $Q_{\mathbf{B}}(\mathbf{B})$ can be factorize into $\prod_{n=1}^N \prod_{l=1}^L Q_b(b_{nl})$. Such a factorization for the postulated posterior distribution of the bits, where the distribution is assumed to be independent over n and l , is commonly known as the mean field approximation. In this paper, we assume $Q_b(b_{nl})$ to be a Bernoulli distribution with parameter λ_{nl} . We postulate the posterior conditional distribution of phase noise to be a multi-variate Gaussian distribution with mean \mathbf{m}_{θ} and covariance \mathbf{S}_{θ} . Thus, we have

$$\begin{aligned} Q(\mathbf{B}, \theta) &= Q_{\mathbf{B}}(\mathbf{B})Q_{\theta}(\theta) \\ &= \left[\prod_{n=1}^N \prod_{l=1}^L (\lambda_{nl})^{\frac{1+b_{nl}}{2}} (1 - \lambda_{nl})^{\frac{1-b_{nl}}{2}} \right] \mathcal{N}(\mathbf{m}_{\theta}, \mathbf{S}_{\theta}) \end{aligned} \quad (12)$$

$$= \left[\prod_{n=1}^N \prod_{l=1}^L \left(\frac{1+\hat{b}_{nl}}{2} \right)^{\frac{1+b_{nl}}{2}} \left(\frac{1-\hat{b}_{nl}}{2} \right)^{\frac{1-b_{nl}}{2}} \right] \mathcal{N}(\mathbf{m}_{\theta}, \mathbf{S}_{\theta}) \quad (13)$$

where, $\hat{b}_{nl} = 2\lambda_{nl} - 1$, is the mean of the postulated posterior distribution $Q_b(b_{nl})$. Having fixed the structure of the postulated posterior distribution, it remains now to compute the parameters of this distribution such that it closely approximates the actual posterior distribution. The variational inference approach introduces the concept of variational free energy as a measure of similarity between two distributions. The variational free energy between the two distributions of interest here is given by

$$F(Q, p) = \int_{\mathbf{B}, \theta} Q(\mathbf{B}, \theta) \log \frac{Q(\mathbf{B}, \theta)}{p(\mathbf{B}, \theta, \mathbf{r})} d\mathbf{B} d\theta. \quad (14)$$

We use the distribution $p(\mathbf{B}, \theta, \mathbf{r})$ instead of $p(\mathbf{B}, \theta | \mathbf{r})$ as they are related simply through a constant of proportionality. It is to be noted that the free energy expression is exactly equal to the Kullback-Leibler divergence between $Q(\mathbf{B}, \theta)$ and $p(\mathbf{B}, \theta | \mathbf{r})$ to within an additive constant. The variational inference approach involves the minimization of the free energy over the parameters of $Q(\mathbf{B}, \theta)$ so that the resulting distribution closely approximates the actual posterior distribution.

The actual posterior distribution can be computed by conditioning over the unknowns as follows:

$$p(\mathbf{B}, \theta, \mathbf{r}) = p(\mathbf{r} | \mathbf{B}, \theta) p(\mathbf{B}) p(\theta) \quad (15)$$

Using (9), and assuming independent Bernoulli distributed priors on bits with the l^{th} bit of the n^{th} symbol having mean μ_{nl} (set μ_{nl} to 0 if no prior information is available) and assuming the prior distribution of PHN to be a multi-variate Gaussian distribution with mean μ_{θ} and

covariance matrix ϕ_θ (if no prior information is available, set mean to $\mathbf{0}$ and covariance to Φ as given in (2)), we can write the posterior distribution as

$$p(\mathbf{B}, \boldsymbol{\theta}, \mathbf{r}) = \mathcal{CN}(\text{diag}(\mathbf{P}\mathbf{F}^H\mathbf{H}f(\mathbf{B}))(1 + j\boldsymbol{\theta}), 2\sigma^2\mathbf{I}) \left[\prod_{n=1}^N \prod_{l=1}^L \left(\frac{1+\mu_{nl}}{2} \right)^{\frac{1+b_{nl}}{2}} \left(\frac{1-\mu_{nl}}{2} \right)^{\frac{1-b_{nl}}{2}} \right] \mathcal{N}(\boldsymbol{\mu}_\theta, \phi_\theta). \quad (16)$$

The next section discusses the computation and minimization of the free energy expression.

IV. THE BIT-LEVEL VARIATIONAL INFERENCE ALGORITHM

A. Free Energy Evaluation

The free energy expression given in (14) can be written as the summation of five terms as shown in (17).

$$\begin{aligned} F(Q, p) = & - \underbrace{\int_{\mathbf{B}} Q(\boldsymbol{\theta}) \log p(\mathbf{B}) d\mathbf{B}}_i - \underbrace{\int_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) d\boldsymbol{\theta}}_{ii} + \underbrace{\int_{\mathbf{B}} Q(\mathbf{B}) \log p(\mathbf{B}) d\mathbf{B}}_{iii} \\ & + \underbrace{\int_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}) \log Q(\boldsymbol{\theta}) d\boldsymbol{\theta}}_{iv} - \underbrace{\int_{\mathbf{B}, \boldsymbol{\theta}} Q(\mathbf{B}) Q(\boldsymbol{\theta}) \log(p(\mathbf{r}|\mathbf{B}, \boldsymbol{\theta})) d\mathbf{B} d\boldsymbol{\theta}}_v \end{aligned} \quad (17)$$

The exact computation of the free energy expression is given in Appendix A and the final expression is presented here. Note that the free energy expression is parameterized by the mean and the covariance matrix of the PHN sequence and the mean value of the posterior estimate of the bits. Note that when treating the unknown \mathbf{B} as a matrix of Bernoulli random variables, we denote $\hat{\mathbf{B}}$ as the matrix of the means (\hat{b}_{nl}) of the corresponding random variables in \mathbf{B} . We define $\hat{\mathbf{B}}_r$, $\hat{\mathbf{B}}_i$, $\hat{\mathbf{b}}_{rj}$ and $\hat{\mathbf{b}}_{rk}$ in a similar fashion.

$$\begin{aligned} F(\mathbf{m}_\theta, \mathbf{S}_\theta, \hat{\mathbf{B}}) = & - \sum_{n=1}^N \sum_{l=1}^L \left[\left(\frac{1+\hat{b}_{nl}}{2} \right) \log \left(\frac{1+\mu_{nl}}{2} \right) + \left(\frac{1-\hat{b}_{nl}}{2} \right) \log \left(\frac{1-\mu_{nl}}{2} \right) \right] + \frac{1}{2} \text{tr}(\phi_\theta^{-1} \mathbf{S}_\theta) + \frac{1}{2} \mathbf{m}_\theta^T \phi_\theta^{-1} \mathbf{m}_\theta \\ & - \boldsymbol{\mu}_\theta^T \phi_\theta^{-1} \mathbf{m}_\theta + \sum_{n=1}^N \sum_{l=1}^L \left[\left(\frac{1+\hat{b}_{nl}}{2} \right) \log \left(\frac{1+\hat{b}_{nl}}{2} \right) + \left(\frac{1-\hat{b}_{nl}}{2} \right) \log \left(\frac{1-\hat{b}_{nl}}{2} \right) \right] - \frac{1}{2} \log |\mathbf{S}_\theta| \\ & + \frac{1}{2\sigma^2} \left[-\mathbf{r}^H \mathbf{Z} f(\hat{\mathbf{B}}) - f(\hat{\mathbf{B}})^H \mathbf{Z}^H \mathbf{r} + f_r(\hat{\mathbf{B}}_r)^H \mathbf{M}_0 f_i(\hat{\mathbf{B}}_i) + f_i(\hat{\mathbf{B}}_i)^H \mathbf{M}_0 f_r(\hat{\mathbf{B}}_r) \right. \\ & \left. + f_r(\hat{\mathbf{B}}_r)^H \mathbf{M}_1 f_r(\hat{\mathbf{B}}_r) + f_i(\hat{\mathbf{B}}_i)^H \mathbf{M}_1 f_i(\hat{\mathbf{B}}_i) + \mathbf{1}^H \mathbf{M}_2 \boldsymbol{\nu}_r + \mathbf{1}^H \mathbf{M}_2 \boldsymbol{\nu}_i \right] \end{aligned} \quad (18)$$

B. Free Energy Minimization

Clearly, closed form expressions for the optimal parameters that minimize the free energy expression cannot be computed. Hence, we adopt a coordinate-descent approach to the minimization, wherein, one parameter is updated while keeping the others constant. Such an approach will converge to a local minimum. The update to each parameter is obtained by computing the gradient of the free energy expression given in (18) w.r.t the parameter and setting it to zero. This leads to three update equations, one each for the mean and covariance matrix of the phase noise and one for updating the mean corresponding to the bits. The update equations are :

$$\mathbf{S}_\theta^T = \left[\phi_\theta^{-1} + \frac{1}{\sigma^2} \mathbf{X}_m \mathbf{X}_m^H + \frac{1}{\sigma^2} \left(\text{diag} \left(\mathbf{F}^H \mathbf{H} \text{diag}(\boldsymbol{\nu}_r + \boldsymbol{\nu}_i - |f(\hat{\mathbf{B}})|^2) \mathbf{H}^H \mathbf{F} \right) \right) \right]^{-1} \quad (19)$$

$$\mathbf{m}_\theta = \mathbf{S}_\theta \left[-\frac{1}{\sigma^2} \text{Im}(\mathbf{r}^H \mathbf{X}_m)^T + \phi_\theta^{-1} \boldsymbol{\mu}_\theta \right] \quad (20)$$

$$\mathbf{t}_{brk} = \mathbf{t}_{\mu rk} + \frac{1}{\sigma^2} \left[\text{diag}(\boldsymbol{\alpha}_k) \Re\{\mathbf{Z}^H \mathbf{r}\} - \text{diag}(\boldsymbol{\delta}_k) \mathbf{M}_1^T \mathbf{1} + \text{diag}(\boldsymbol{\alpha}_k) \left(j \Im\{\mathbf{M}_0\} f_i(\hat{\mathbf{B}}_i) - \Re\{\mathbf{M}_2^T\} f_r(\hat{\mathbf{B}}_r) \right) \right] \quad (21)$$

$$\mathbf{t}_{bik} = \mathbf{t}_{\mu ik} + \frac{1}{\sigma^2} \left[\text{diag}(\boldsymbol{\beta}_k) \Im\{\mathbf{Z}^H \mathbf{r}\} - \text{diag}(\boldsymbol{\Omega}_k) \mathbf{M}_1^T \mathbf{1} + \text{diag}(\boldsymbol{\beta}_k) \left(\Im\{\mathbf{M}_0\} f_r(\hat{\mathbf{B}}_r) + j \Re\{\mathbf{M}_2^T\} f_i(\hat{\mathbf{B}}_i) \right) \right] \quad (22)$$

In the above equations, \mathbf{t}_{brk} and \mathbf{t}_{bik} represent $\tanh^{-1}(\hat{\mathbf{b}}_{rk})$ and $\tanh^{-1}(\hat{\mathbf{b}}_{ik})$ (computed element wise) respectively, while $\mathbf{t}_{\mu rk}$ and $\mathbf{t}_{\mu ik}$ represent $\tanh^{-1}(\boldsymbol{\mu}_{rk})$ and $\tanh^{-1}(\boldsymbol{\mu}_{ik})$ respectively, with $\boldsymbol{\mu}_{rk}$ and $\boldsymbol{\mu}_{ik}$ defined analogous to $\hat{\mathbf{b}}_{rk}$ and $\hat{\mathbf{b}}_{ik}$ but with prior means. Further, the derivatives $\frac{\partial f_r(\hat{\mathbf{B}}_r)}{\partial \hat{\mathbf{b}}_{rk}}$ and $\frac{\partial f_i(\hat{\mathbf{B}}_i)}{\partial \hat{\mathbf{b}}_{ik}}$ are denoted as $\text{diag}(\boldsymbol{\alpha}_k)$ and $\text{diag}(\boldsymbol{\beta}_k)$. The derivatives $\frac{\partial \boldsymbol{\nu}_r}{\partial \hat{\mathbf{b}}_{rk}}$ and $\frac{\partial \boldsymbol{\nu}_i}{\partial \hat{\mathbf{b}}_{ik}}$ are denoted as $\text{diag}(\boldsymbol{\delta}_k)$ and $\text{diag}(\boldsymbol{\Omega}_k)$. Detailed derivation of the update equations is given in Appendix B.

C. The Variational Inference Algorithm

The pseudo code given in the next page gives the steps involved in the VI based soft bit detection algorithm. It was noted that because of the the coordinate descent approach to free energy minimization, the algorithm converges to a local minimum and not the global minimum. This was particularly found to be an issue in scenarios of high CPE. To overcome this, we use the second term of (17) (denoted as F_2 henceforth) as an indicator of convergence to the right PHN sequence. Whenever F_2 is greater than a certain threshold, it indicates that the estimated

Algorithm 1 Bit level Variational Inference Algorithm

```

1: Initialize.  $\mu_{nl} \leftarrow 0$  or given priors;  $\boldsymbol{\mu}_\theta \leftarrow \mathbf{0}$ ;  $\boldsymbol{\phi}_\theta \leftarrow \bar{\boldsymbol{\Phi}}$ ;  $m_{nl} \leftarrow 0$ ;  $\mathbf{m}_\theta \leftarrow \mathbf{0}$ ;  $\mathbf{S}_\theta \leftarrow \mathbf{0}$ ;
2: Compute  $\mathbf{t}_{\mu rk}$  and  $\mathbf{t}_{\mu ik}$  for  $k \in \{1, 2, \dots, \frac{L}{2}\}$ .
3: for  $s=1:num\_iter$  do
4:   Compute  $\boldsymbol{\nu}_r, \boldsymbol{\nu}_i, \boldsymbol{\alpha}_k, \boldsymbol{\delta}_k, \mathbf{X}_m$ .
5:   Update  $\mathbf{S}_\theta$  using (19).
6:   Update  $\mathbf{m}_\theta$  using (20).
7:   for  $k=L/2:1$  do
8:     Compute  $\mathbf{t}_{brk}$  and  $\mathbf{t}_{bik}$  using (21) and (22).
9:     for  $p=1:L/2$  do
10:      Update  $\hat{\mathbf{B}}, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_p, \boldsymbol{\delta}_p, \boldsymbol{\Omega}_p$ .
11:    end for
12:  end for
13: end for
14: Compute  $F_2$  in (17).
15: if  $F_2 > threshold$  then
16:   Ignore PHN and detect symbols.
17: else
18:   Return  $2(\mathbf{t}_{brk} - \mathbf{t}_{\mu rk})$  &  $2(\mathbf{t}_{bik} - \mathbf{t}_{\mu ik})$  for  $k \in \{1, 2, \dots, \frac{L}{2}\}$ .
19: end if

```

PHN sequence does not correspond to the expected statistics and suggests that the algorithm has converged to a wrong sequence. In such scenarios, we ignore the output of the algorithm and compute the soft bit estimates ignoring PHN.

V. SIMULATION RESULTS

To test the performance of the suggested algorithm we set up the following simulation. We simulated a link using 64-QAM constellation and OFDM with 64 sub-carriers over a frequency selective channel. The channel was assumed to be a Rayleigh multipath fading channel with 10 taps and an exponential power delay profile. The sampling rate was set to 20 MHz, which in turn implies a sub-carrier spacing of 312.5 kHz. The oscillator bandwidth was set to 100 kHz

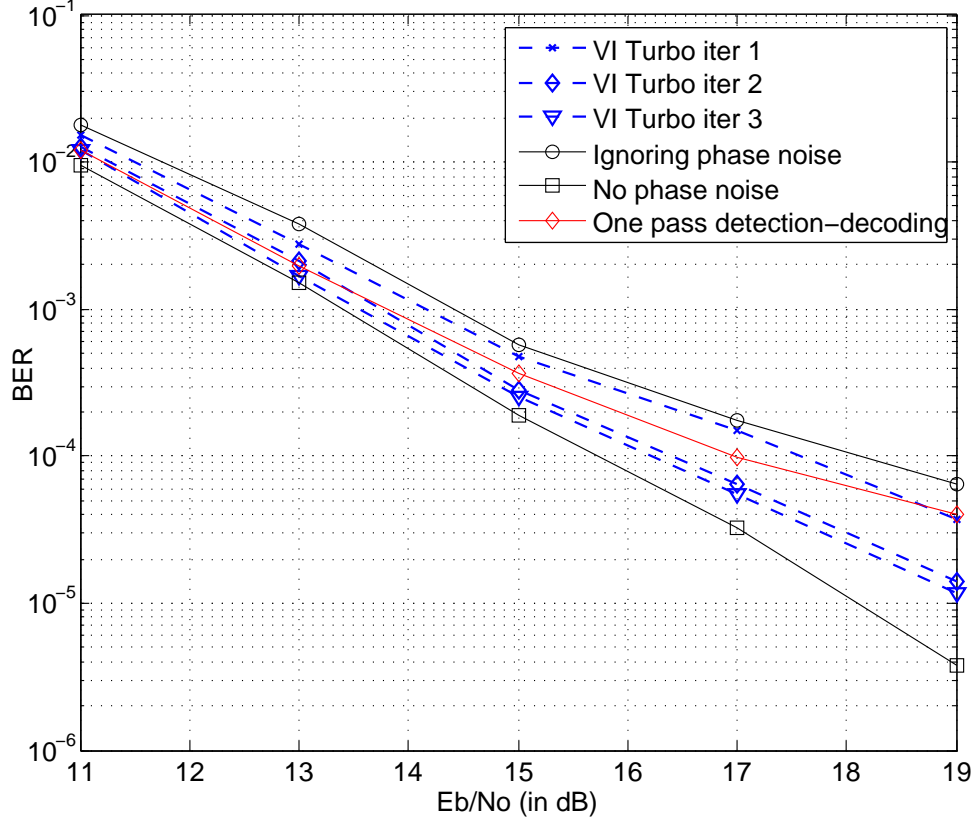


Fig. 1: BER plot of the turbo receiver.

and the standard deviation σ_θ was set to 3° . The MATLAB code presented in [2] was used to generate the PHN sequences.

For the turbo receiver setup, an outer LDPC code of rate $3/4$ and length 2304 was used. On the transmit side, the message bits were encoded, interleaved and mapped to symbols from the 64-QAM constellation. The length of the outer code was chosen so as to span 6 OFDM symbols. On the receiver side, after the reception of the 6 OFDM symbols, each was passed through the bit level detection algorithm and the extrinsic information so obtained was passed to the soft decoder after deinterleaving. To begin with, the detector was initialized to uniform priors; in subsequent iterations, the extrinsic information obtained from the decoder was used as priors. For this particular simulation, we used 3 outer iterations with 6 iterations for the decoder and 5 iterations for the detector. Fig. 1 presents the results of this simulation and shows a significant improvement in performance with every subsequent iteration of the turbo receiver. It can also be seen that the turbo receiver clearly out-performs a one pass receiver where there is no loop

back between the detector and decoder. The receiver performance is compared against scenarios where there is no PHN and where PHN is completely ignored. For both these cases, and for the one pass setup, the FEC decoder was run for 18 iterations.

VI. CONCLUSION

In this paper we highlighted the consequences of high CPE and its effect on decision directed algorithms. To overcome this effect, we developed a bit-level soft detection algorithm using the VI approach and used this algorithm to build a turbo receiver. Through simulations we have established the performance gains that can be achieved using the turbo receiver.

APPENDIX A

EVALUATION OF THE FREE ENERGY EXPRESSION

As shown in (17), the free energy expression can be written as the sum of five terms. Since we have adopted a discrete distribution for $Q_{\mathbf{B}}(\mathbf{B})$, it is straightforward to compute the first and third terms. The second and fourth term together constitute the Kullback-Leibler divergence between two multi-variate Gaussian distributions and can be computed easily. We discuss the computation of the fifth term here. Now,

$$-2\sigma^2(v) = \text{const.} + \int_{\mathbf{B}, \boldsymbol{\theta}} \left[Q(\mathbf{B}) Q(\boldsymbol{\theta}) (\mathbf{r} - \mathbf{P}\mathbf{F}^H \mathbf{H} f(\mathbf{B}))^H (\mathbf{r} - \mathbf{P}\mathbf{F}^H \mathbf{H} f(\mathbf{B})) \right] d\mathbf{B} d\boldsymbol{\theta} \quad (23)$$

Proceeding exactly as in [7], we can simplify (23) to

$$-2\sigma^2(v) = \text{const.} + \int_{\mathbf{B}} Q(\mathbf{B}) \left[f(\mathbf{B})^H \mathbf{Z}^H \mathbf{Z} f(\mathbf{B}) - \mathbf{r}^H \mathbf{Z} f(\mathbf{B}) - f(\mathbf{B})^H \mathbf{Z}^H \mathbf{r} + f(\mathbf{B})^H \boldsymbol{\Psi} f(\mathbf{B}) \right] d\mathbf{B} \quad (24)$$

Here, we have defined $\mathbf{Z} = \text{diag}(\mathbf{1} + j\mathbf{m}_{\theta}) \mathbf{F}^H \mathbf{H}$ and $\boldsymbol{\Psi} = \mathbf{H}^H \mathbf{F} \text{diag}(\mathbf{S}_{\theta}) \mathbf{F}^H \mathbf{H}$. Using Lemmas 5 and 6 from [13], we can simplify this to

$$\begin{aligned} -2\sigma^2(v) = & \text{const.} - \mathbf{r}^H \mathbf{Z} f(\hat{\mathbf{B}}) - f(\hat{\mathbf{B}})^H \mathbf{Z}^H \mathbf{r} + \mathbf{1}^H (\mathbf{M}_1) (\boldsymbol{\nu}_r + \boldsymbol{\nu}_r) + f_r(\hat{\mathbf{B}}_r)^H (\mathbf{M}_0) f_i(\hat{\mathbf{B}}_i) \\ & + f_i(\hat{\mathbf{B}}_i)^H (\mathbf{M}_0) f_r(\hat{\mathbf{B}}_r) + f_r(\hat{\mathbf{B}}_r)^H (\mathbf{M}_2) f_r(\hat{\mathbf{B}}_r) + f_i(\hat{\mathbf{B}}_i)^H (\mathbf{M}_2) f_i(\hat{\mathbf{B}}_i). \end{aligned} \quad (25)$$

In the equation above, we have used the following definitions.

$$\mathbf{M}_0 = (\Psi + \mathbf{Z}^H \mathbf{Z}) \quad (26)$$

$$\mathbf{M}_1 = (\Psi + \mathbf{Z}^H \mathbf{Z} - \text{diag}(\Psi + \mathbf{Z}^H \mathbf{Z})) \quad (27)$$

$$\mathbf{M}_2 = \text{diag}(\Psi + \mathbf{Z}^H \mathbf{Z}) \quad (28)$$

$$\nu_r = \sum_{0 < i \leq j < (L/2)} 2^{i+j} \prod_{p=i}^{p=j} \hat{\mathbf{b}}_p + \sum_{i=1}^{L/2} 2^{2(i-1)} \quad (29)$$

$$\nu_i = \sum_{0 < i \leq j < (L/2)} 2^{i+j} \prod_{p=i}^{p=j} \hat{\mathbf{b}}_{p+(L/2)} + \sum_{i=1}^{L/2} 2^{2(i-1)} \quad (30)$$

APPENDIX B

COMPUTING THE GRADIENT

A. Gradient w.r.t \mathbf{S}_θ

Using (18) and (14), we can write

$$\frac{\partial \mathbf{F}}{\partial \mathbf{S}_\theta^{-1}} = \frac{\partial(-i - ii + iii + iv - v)}{\partial \mathbf{S}_\theta^{-1}} \quad (31)$$

Using results from [14], one can compute the terms above to be :

$$\frac{\partial(i)}{\partial \mathbf{S}_\theta^{-1}} = 0; \quad \frac{\partial(ii)}{\partial \mathbf{S}_\theta^{-1}} = \frac{1}{2} \mathbf{S}_\theta^T \phi_\theta^{-1} \mathbf{S}_\theta^T; \quad \frac{\partial(iii)}{\partial \mathbf{S}_\theta^{-1}} = 0; \quad \frac{\partial(iv)}{\partial \mathbf{S}_\theta^{-1}} = \frac{1}{2} \mathbf{S}_\theta^T \quad (32)$$

To compute $\frac{\partial(v)}{\partial \mathbf{S}_\theta^{-1}}$, note the following two results [7] :

$$\begin{aligned} \text{(a)} \quad \frac{\partial(\mathbf{x}^H \text{diag}(\mathbf{S}_\theta) \mathbf{y})}{\partial \mathbf{S}_\theta^{-1}} &= \frac{\partial(\text{tr}(\text{diag}(\mathbf{x}^H) \mathbf{S}_\theta \text{diag}(\mathbf{y})))}{\partial \mathbf{S}_\theta^{-1}} \\ &= \frac{\partial(\text{tr}(\text{diag}(\mathbf{y}) \text{diag}(\mathbf{x}^H) \mathbf{S}_\theta))}{\partial \mathbf{S}_\theta^{-1}} \\ &= -\mathbf{S}_\theta^T (\text{diag}(\mathbf{y}) \text{diag}(\mathbf{x}^H)) \mathbf{S}_\theta \end{aligned} \quad (33)$$

$$\begin{aligned} \text{(b)} \quad \frac{\partial(\text{tr}(\mathbf{X} \text{diag}(\mathbf{S}_\theta)))}{\partial \mathbf{S}_\theta^{-1}} &= \frac{\partial(\text{tr}(\text{diag}(\mathbf{X}) \mathbf{S}_\theta))}{\partial \mathbf{S}_\theta^{-1}} \\ &= -\mathbf{S}_\theta^T (\text{diag}(\mathbf{X})) \mathbf{S}_\theta^T \end{aligned} \quad (34)$$

In the above equations, by $\text{diag}(\mathbf{S}_\theta)$ we mean the diagonal matrix formed using the diagonal entries of the matrix \mathbf{S}_θ . Using (33) and (34) we can compute $\frac{\partial(v)}{\partial \mathbf{S}_\theta^{-1}}$ to be

$$\frac{\partial(v)}{\partial \mathbf{S}_\theta^{-1}} = \frac{-1}{2\sigma^2} \left[-\mathbf{S}_\theta^T (\mathbf{X}_m \mathbf{X}_m^H) \mathbf{S}_\theta^T - \mathbf{S}_\theta^T \text{diag}(\mathbf{F}^H \mathbf{H} \text{diag}(\nu_r + \nu_i - |f(\hat{\mathbf{B}})|^2) \mathbf{H}^H \mathbf{F}) \mathbf{S}_\theta^T \right] \quad (35)$$

Using (32) and (35) it is straightforward to compute (19).

B. Gradient w.r.t \mathbf{m}_θ

Using (17) and (18), we can write

$$\frac{\partial \mathbf{F}}{\partial \mathbf{m}_\theta} = \frac{\partial(-i - ii + iii + iv - v)}{\partial \mathbf{m}_\theta}. \quad (36)$$

Using results from [14], one can compute the terms above to be :

$$\frac{\partial(i)}{\partial \mathbf{m}_\theta} = 0; \quad \frac{\partial(ii)}{\partial \mathbf{m}_\theta} = \phi_\theta^{-T} \boldsymbol{\mu}_\theta - \phi_\theta^{-1} \mathbf{m}_\theta; \quad \frac{\partial(iii)}{\partial \mathbf{m}_\theta} = 0; \quad \frac{\partial(iv)}{\partial \mathbf{m}_\theta} = 0; \quad (37)$$

To compute $\frac{\partial(v)}{\partial \mathbf{m}_\theta}$, note the following two results. For any two given vectors \mathbf{x} and \mathbf{y} , using results from [14], we have :

(a)

$$\begin{aligned} \frac{\partial (\mathbf{x}^H (\mathbf{I} + j \text{diag}(\mathbf{m}_\theta))^H (\mathbf{I} + j \text{diag}(\mathbf{m}_\theta)) \mathbf{y})}{\partial \mathbf{m}_\theta} &= \frac{\partial ((\mathbf{1} + j \mathbf{m}_\theta)^H \text{diag}(\mathbf{x}^H) \text{diag}(\mathbf{y}) (\mathbf{1} + j \mathbf{m}_\theta))}{\partial \mathbf{m}_\theta} \\ &= -2 \text{diag}(\mathbf{x}^H) \text{diag}(\mathbf{y}) \mathbf{m}_\theta \end{aligned} \quad (38)$$

(b)

$$\begin{aligned} \frac{\partial (\mathbf{x}^H \text{diag}(\mathbf{Z}^H \mathbf{Z}) \mathbf{y})}{\partial \mathbf{m}_\theta} &= \frac{\partial (\text{tr}(\text{diag}(\mathbf{x}^H) \mathbf{Z}^H \mathbf{Z} \text{diag}(\mathbf{y})))}{\partial \mathbf{m}_\theta} \\ &= \frac{\partial ((\mathbf{1} + j \mathbf{m}_\theta)^T \text{diag}(\mathbf{F}^H \mathbf{H} \text{diag}(\mathbf{y}) \text{diag}(\mathbf{x}^H) \mathbf{H}^H \mathbf{F}) (\mathbf{1} - j \mathbf{m}_\theta))}{\partial \mathbf{m}_\theta} \\ &= 2 \text{diag}(\mathbf{F}^H \mathbf{H} \text{diag}(\mathbf{y}) \text{diag}(\mathbf{x}^H) \mathbf{H}^H \mathbf{F}) \mathbf{m}_\theta \end{aligned} \quad (39)$$

Using (38) and (39), we can compute $\frac{\partial(v)}{\partial \mathbf{m}_\theta}$ to be :

$$\frac{\partial(v)}{\partial \mathbf{m}_\theta} = \frac{-1}{\sigma^2} \left[\mathbf{X}_m^H \mathbf{X}_m \mathbf{m}_\theta + \Im(\mathbf{r}^H \mathbf{X}_m)^T + \text{diag}(\mathbf{F}^H \mathbf{H} \text{diag}(\nu_r + \nu_i - |f(\hat{\mathbf{B}})|^2)) \mathbf{m}_\theta \right] \quad (40)$$

Using (37), (40) and the update equation for \mathbf{S}_θ given in (19), we can compute the update equation for \mathbf{m}_θ to be as given in (20).

C. Gradient w.r.t $\hat{\mathbf{b}}_{rk}$ and $\hat{\mathbf{b}}_{ik}$

To compute $\frac{\partial F}{\partial \hat{\mathbf{b}}_{rk}}$, consider the term (i) from (17):

$$(i) = \sum_{n=1}^N \sum_{l=1}^L \left[\left(\frac{1 + \hat{b}_{nl}}{2} \right) \log \left(\frac{1 + \mu_{nl}}{2} \right) + \left(\frac{1 - \hat{b}_{nl}}{2} \right) \log \left(\frac{1 - \mu_{nl}}{2} \right) \right] \quad (41)$$

Differentiating (41) w.r.t to \hat{b}_{nl} gives $\tanh^{-1}(\mu_{nl})$. Thus, if we denote $\frac{\partial(i)}{\partial \hat{\mathbf{b}}_{rk}}$ as $\mathbf{t}_{\mu rk}$ and $\frac{\partial(i)}{\partial \hat{\mathbf{b}}_{ik}}$ as $\mathbf{t}_{\mu ik}$ for $k \in \{1, 2, \dots, \frac{L}{2}\}$, we have $\mathbf{t}_{\mu rk} = \tanh^{-1}(\mu_{rk})$ and $\mathbf{t}_{\mu ik} = \tanh^{-1}(\mu_{ik})$. Similarly, differentiating term (ii) (given in (42)) w.r.t to \hat{b}_{nl} gives $\tanh^{-1}(\hat{b}_{nl})$. Denoting $\frac{\partial(ii)}{\partial \hat{\mathbf{b}}_{rk}}$ as $\mathbf{t}_{b rk}$ and $\frac{\partial(ii)}{\partial \hat{\mathbf{b}}_{ik}}$ as $\mathbf{t}_{b ik}$, we have $\mathbf{t}_{b rk} = \tanh^{-1}(\hat{\mathbf{b}}_{rk})$ and $\mathbf{t}_{b ik} = \tanh^{-1}(\hat{\mathbf{b}}_{ik})$.

$$(ii) = \sum_{n=1}^N \sum_{l=1}^L \left[\left(\frac{1 + \hat{b}_{nl}}{2} \right) \log \left(\frac{1 + \hat{b}_{nl}}{2} \right) + \left(\frac{1 - \hat{b}_{nl}}{2} \right) \log \left(\frac{1 - \hat{b}_{nl}}{2} \right) \right] \quad (42)$$

Clearly, the terms (iii) and (iv) of 17 are independent of $\hat{\mathbf{b}}_{rk}$ and $\hat{\mathbf{b}}_{ik}$. It remains to compute the derivative of the fifth term. To compute $\frac{\partial(v)}{\partial \hat{\mathbf{b}}_{rk}}$ and $\frac{\partial(v)}{\partial \hat{\mathbf{b}}_{ik}}$, we note the following :

(a) From [13],

$$\frac{\partial f_r(\hat{\mathbf{B}}_r)}{\partial \hat{\mathbf{b}}_{rk}} = \text{diag}(\alpha_k) \text{ where, } \alpha_k = \mathbf{I}(k = L/2)2^{k-1} \cdot \mathbf{1} + \sum_{l=1}^k 2^{l-1} \prod_{p=l, p \neq k}^{L/2} \hat{\mathbf{b}}_{rp} \quad (43)$$

(b) From [13],

$$\frac{\partial f_i(\hat{\mathbf{B}}_i)}{\partial \hat{\mathbf{b}}_{ik}} = \text{diag}(\beta_k) \text{ where, } \beta_k = j \cdot \left(\mathbf{I}(k = L/2)2^{k-1} \cdot \mathbf{1} + \sum_{l=1}^k 2^{l-1} \prod_{p=l, p \neq k}^{L/2} \hat{\mathbf{b}}_{ip} \right) \quad (44)$$

(c) For any given matrix \mathbf{M} with complex entries,

$$\frac{\partial f_r(\hat{\mathbf{B}}_r)^H \mathbf{M} f_r(\hat{\mathbf{B}}_r)}{\partial \hat{\mathbf{b}}_{rk}} = \frac{\partial f_r(\hat{\mathbf{B}}_r)^H \mathbf{M} f_r(\hat{\mathbf{B}}_r)}{\partial f_r(\hat{\mathbf{B}}_r)} \frac{\partial f_r(\hat{\mathbf{B}}_r)}{\partial \hat{\mathbf{b}}_{rk}} = 2 \left[\Re \{ f_r(\hat{\mathbf{B}}_r)^H \mathbf{M} \} \text{diag}(\alpha_{rk}) \right]^T \quad (45)$$

(d) For any given matrix \mathbf{M} with complex entries,

$$\frac{\partial f_i(\hat{\mathbf{B}}_i)^H \mathbf{M} f_i(\hat{\mathbf{B}}_i)}{\partial \hat{\mathbf{b}}_{ik}} = \frac{\partial f_i(\hat{\mathbf{B}}_i)^H \mathbf{M} f_i(\hat{\mathbf{B}}_i)}{\partial f_i(\hat{\mathbf{B}}_i)} \frac{\partial f_i(\hat{\mathbf{B}}_i)}{\partial \hat{\mathbf{b}}_{ik}} = 2 \left[\Im \{ f_i(\hat{\mathbf{B}}_i)^H \mathbf{M} \} \text{diag}(\beta_{ik}) \right]^T \quad (46)$$

(e) From [13],

$$\frac{\partial \nu_r}{\partial \hat{\mathbf{b}}_{rk}} = \text{diag} \left(\mathbf{I}(0 < k < L/2) \cdot 2^{2k} \mathbf{1} + \sum_{\substack{0 < i \leq k \leq j < L/2 \\ i \neq j}} \prod_{\substack{p=i \\ p \neq k}}^j \hat{b}_{rp} \right) \triangleq \text{diag}(\delta_k) \quad (47)$$

(f) From [13],

$$\frac{\partial \nu_i}{\partial \hat{\mathbf{b}}_{ik}} = \text{diag} \left(\mathbf{I}(0 < k < L/2) \cdot 2^{2k} \mathbf{1} + \sum_{\substack{0 < i \leq k \leq j < L/2 \\ i \neq j}} \prod_{\substack{p=i \\ p \neq k}}^j \hat{b}_{ip} \right) \triangleq \text{diag}(\Omega_k) \quad (48)$$

(g) For any given matrix \mathbf{M} ,

$$\frac{\partial(\mathbf{1}^T \mathbf{M} \nu_r)}{\partial \hat{\mathbf{b}}_{rk}} = \mathbf{M}^T \mathbf{1} \text{diag}(\delta_k) \quad (49)$$

(h) For any given matrix \mathbf{M} ,

$$\frac{\partial(\mathbf{1}^T \mathbf{M} \nu_i)}{\partial \hat{\mathbf{b}}_{ik}} = \mathbf{M}^T \mathbf{1} \text{diag}(\Omega_k) \quad (50)$$

In (47) and (48), $\mathbf{I}(\cdot)$ represents the indicator function which is 1 if the argument is 0 and 0 otherwise. Using all of the results from equations (43) to (50), we can compute the gradient to be :

$$\begin{aligned} \frac{\partial F}{\partial \hat{\mathbf{b}}_{rk}} = & \mathbf{t}_{\mu rk} - \mathbf{t}_{brk} + \frac{1}{\sigma^2} \left[\text{diag}(\alpha_k) \left(\Re\{\mathbf{Z}^H \mathbf{r}\} - \Re\{\mathbf{M}_2^T\} f_r(\hat{\mathbf{B}}_r) \right) \right. \\ & \left. + j \text{diag}(\alpha_k) \Im\{\mathbf{M}_0\} f_i(\hat{\mathbf{B}}_i) - \text{diag}(\delta_k) \mathbf{M}_1^T \mathbf{1} \right] \end{aligned} \quad (51)$$

$$\begin{aligned} \frac{\partial F}{\partial \hat{\mathbf{b}}_{ik}} = & \mathbf{t}_{\mu ik} - \mathbf{t}_{bik} + \frac{1}{\sigma^2} \left[\text{diag}(\beta_k) \left(\Im\{\mathbf{Z}^H \mathbf{r}\} + \Im\{\mathbf{M}_0\} f_r(\hat{\mathbf{B}}_r) \right) \right. \\ & \left. + j \cdot \text{diag}(\beta_k) \Re\{\mathbf{M}_2^T\} f_i(\hat{\mathbf{B}}_i) - \text{diag}(\Omega_k) \mathbf{M}_1^T \mathbf{1} \right] \end{aligned} \quad (52)$$

REFERENCES

- [1] D. Petrovic, W. Rave, and G. Fettweis, "Performance degradation of coded-OFDM due to phase noise," in *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, vol. 2, April 2003, pp. 1168–1172 vol.2.
- [2] M. Webster and M. Seals. Suggested phase noise model for 802.11hrb. [Online]. Available: <https://mentor.ieee.org/802.11/dcn/00/11-00-0296-01-00sb-suggested-phase-noise-model-for-802-11-hrb.ppt>
- [3] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *Communications, IEEE Transactions on*, vol. 43, no. 234, pp. 191–193, Feb/Mar/Apr 1995.
- [4] D. Petrovic, W. Rave, and G. Fettweis, "Phase noise influence on bit error rate, cut-off rate and capacity of M-QAM OFDM signaling," in *In Proc. Intl. OFDM Workshop (InOWo)02*, 2002.
- [5] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: consequences and solutions," *Communications, IEEE Transactions on*, vol. 52, no. 11, pp. 1988–1996, Nov. 2004.
- [6] —, "Performance analysis on the effect of phase noise in OFDM systems," in *Spread Spectrum Techniques and Applications, 2002 IEEE Seventh International Symposium on*, vol. 1, 2002, pp. 133–138 vol.1.

- [7] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1862–1874, May 2007.
- [8] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte Carlo methods for channel, phase noise, and frequency offset estimation with unknown noise variances in OFDM systems," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3613–3626, Aug. 2008.
- [9] Y. Gong and X. Hong, "A new algorithm for OFDM joint data detection and phase noise cancellation," in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 636–640.
- [10] W. Rave, D. Petrovic, and G. Fettweis, "Iterative correction of phase noise in multicarrier modulation," in *In Proc. of the 9th International OFDM Workshop InOWo*, 2004.
- [11] A. Barbieri, G. Colavolpe, and G. Caire, "Joint iterative detection and decoding in the presence of phase noise and frequency offset," *Communications, IEEE Transactions on*, vol. 55, no. 1, pp. 171–179, Jan. 2007.
- [12] D. D. Lin, R. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *Signal Processing, IEEE Transactions on*, vol. 54, no. 9, pp. 3542–3554, Sept. 2006.
- [13] D. D. Lin and T. J. Lim, "Bit-level equalization and soft detection for Gray-coded multilevel modulation," *Information Theory, IEEE Transactions on*, vol. 54, no. 10, pp. 4731–4742, Oct. 2008.
- [14] M. Brookes. Matrix reference manual. [Online]. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html#Intro>